*The Canadian Journal of Statistics*
Vol. 21, No. 2, 1993, Pages 209–222
*La Revue Canadienne de Statistique*



# Some nonparametric methods for changepoint problems

Vera R. EASTWOOD*

*Acadia University*

*Key words and phrases*: Projections of $U$-statistics, limiting distributions, distributions of Gaussian functionals.
*AMS 1985 subject classifications*: Primary 60G15; secondary 62G10.

## ABSTRACT

A general model for changepoint problems is discussed from a nonparametric viewpoint. The test statistics introduced are based on Cramér–von Mises functionals of certain processes and are shown to converge in distribution to corresponding Gaussian functionals (under the assumption of no change in distribution, $\mathcal{H}_0$). We also demonstrate how the distribution of the limiting Gaussian functionals may be tabulated. Finally, properties of the tests under the alternative hypothesis of exactly one changepoint occurring are studied, and some examples are given.

## RÉSUMÉ

Cet article examine d'un point de vue non-paramétrique un modèle général pour les problèmes d'identification du moment de changement de loi. Les fonctions des observations à tester sont fondées sur des fonctionnels de type Cramér-von Mises de certains procédés et il est démontré que ces fonctions des observations convergent en distribution vers les fonctionnels de Gauss correspondants (sous l'hypothèse d'aucun changement en distribution, $\mathcal{H}_0$). Nous démontrons également comment tabuler la distribution des fonctionnels limites de Gauss. Enfin, nous étudions des propriétés des tests sous l'hypothèse alternative d'exactement un moment de changement de loi et donnons quelques exemples.

## 1. INTRODUCTION

The paper presented here is partially based on the author's Ph.D. thesis (cf. Huse 1988), which discusses a number of changepoint problems in an AMOC (at most one changepoint) setting. The topics examined there include a sequential approach (see also Huse 1989) as well as fixed-sample-size methods. The latter are applied using Kolmogorov-Smirnov and Cramér–von Mises test statistics as well as the Kendall-Kendall pontogram technique [cf. also Huse-Eastwood (1989) and Eastwood (1990)]. In this article we restrict our attention to the Cramér–von Mises setting. For a more general introduction to nonparametric methods for changepoint problems we refer to Csörgö and Horváth (1988b) and to Wolfe and Schechtman (1984).

The hypotheses under consideration here can be stated as follows:

$\mathcal{H}_0$: $X_1, X_2, \ldots, X_n$ are independent with cumulative distribution function $F$,

versus

*Co-winner of the Pierre Robillard Award in 1989. Research supported in part by an NSERC operating grant.

$\mathcal{H}_A$: there exists $1 \leq k < n$ such that $X_1, \ldots, X_k$ have distribution $F$, that $X_{k+1}, \ldots, X_n$ have distribution $G$, and that $F \neq G$.

The only assumption on the distributions $F$ (and $G$) involved is that they are nondegenerate.

In this general setup we wish to consider test statistics that are based on Cramér–von Mises (i.e., $L_2$-norm) functionals of

$$U_k = \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j) - k(n-k)\theta, \qquad 0 \leq k \leq n, \tag{1.1}$$

where $h$ is either an antisymmetric or a symmetric real-valued function (kernel) and where $\theta = \mathcal{E}\, h(X_1, X_2)$ (i.e., $\theta = 0$ in case $h$ is antisymmetric).

For a given kernel $h$ and any (asymptotically) consistent estimator $\hat{\theta}$ of the parameter $\theta$, the functionals under consideration are of the form

$$\frac{n^{-3}}{\tilde{\sigma}^2} \int_0^1 \frac{U^2_{[(n+1)t]}}{g(t)}\, dt \tag{1.2}$$

for suitable weight functions $g$. Here $\tilde{\sigma}^2$ is the variance of the projection $\tilde{h}$ of the kernel $h$ as defined in Section 2.

The change in distribution of course occurs (if it does at all) at unknown time $k_0$. This implies that under the alternative hypothesis, (1.1) cannot be modelled using two-sample results, since the true changepoint $k_0$ is unknown. However, under the null hypothesis of no change, one-sample $U$-statistics can be related to $U_k$, as will be shown in Section 2. The idea will be that under the assumption of no change in distribution we will first represent $U_k$ as a linear combination of nondegenerate (provided that $h$ is a nondegenerate kernel) one-sample $U$-statistics. Next we will approximate (via the projection principle) each $U$-statistic involved by an appropriate sum of independent identically distributed random variables [see (2.6)]. Then we will employ Donsker's Theorem as in Csörgő and Révész (1981, p. 90) to find the limiting distribution of the Cramér–von Mises functionals (1.2). It will be established that these limits do not in any way depend on either the kernel function $h$ or the initial distribution $F$. Consequently, we can derive approximate cutoff points from those limits, and we only need to tabulate values twice (antisymmetric and symmetric cases separately), regardless of whether we wish to test for a shift in mean or a change in variance or any other change in distribution.

The remaining problem then is whether we can find a general enough method to actually compute these percentage points (especially for complicated weighted versions of the test statistics). The question of weights is discussed along with the derivation of the limiting distributions in Section 2, and methods for the tabulation of the cutoff points are presented in Section 3. Section 4 contains considerations under the alternative hypothesis of exactly one changepoint occurring at unknown time $k$, and also some concluding remarks.

We close this introductory section with some examples of tests involving the variables $U_k$:

(1) *Antisymmetric kernel $h$: Wilcoxon rank-sum statistic.* Let $X_1, \ldots, X_n$ denote a random sample from a continuous distribution $F$. Define $h(x,y) = \mathrm{sgn}(x-y) = 1, 0, -1$ if $x > y, x = y, x < y$ respectively. Let $R_1, R_2, \ldots, R_k$ denote the ranks of the first $k$

observations $X_1, \ldots, X_k$ in the complete sample of $n$ observations. Then we obtain [cf. Pettitt (1979) without proof and Huse (1988) with proof]

$$U_k = \sum_{i=1}^{k} \sum_{j=k+1}^{n} \operatorname{sgn}(X_i - X_j) = 2W_k - k(n+1) \qquad \text{for} \quad 0 \leq k \leq n,$$

where $W_k = \sum_{j=1}^{k} R_j$ is the Wilcoxon rank-sum statistic. Under the assumption of no change in distribution ($\mathcal{H}_0$) we see that

$$\theta = \mathcal{E} h(X_1, X_2) = \mathcal{E} \operatorname{sgn}(X_1 - X_2) = 0 \quad \text{and} \quad \mathcal{V}ar\, \tilde{h}^2(X_1) = \tfrac{1}{3} > 0,$$

where $\tilde{h}(t) = \mathcal{E} h(t, X_1)$ is the projection of the kernel $h$. Consequently, all methods presented in the next sections apply here.

(2) *Symmetric kernel h: change-in-variance problems.* Take $\theta = \sigma^2 = \sigma^2(F)$, where $F$ is the cumulative distribution function of the random sample $X_1, \ldots, X_n$, and let $h(x, y) = \frac{1}{2}(x - y)^2$. Then the corresponding $U$-statistic is

$$\begin{aligned} U(X_1, X_2, \ldots, X_n) &= \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} h(X_i, X_j) \\ &= \frac{1}{n(n-1)} \sum_{1 \leq i < j \leq n} (X_i - X_j)^2 \\ &= \frac{1}{n-1} \left( \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right) = S^2 \end{aligned}$$

(cf. also Serfling 1980). Furthermore,

$$\begin{aligned} U_k &= \frac{1}{2} \sum_{i=1}^{k} \sum_{j=k+1}^{n} (X_i - X_j)^2 - k(n-k)\sigma^2 \\ &= \binom{n}{2} U(X_1, \ldots, X_n) - \binom{k}{2} U(X_1, \ldots, X_k) \\ &\quad - \binom{n-k}{2} U(X_{k+1}, \ldots, X_n) - k(n-k)\sigma^2, \end{aligned}$$

i.e., this special $U_k$ is indeed a linear combination of nondegenerate $U$-statistics, as claimed in the introduction. Utilizing a sequence of consistent estimators $\sigma_n^2$ for $\sigma^2 > 0$ and Slutzky's Theorem, we see again that the methods discussed in the following apply here. Further typical choices of symmetric kernels include $h(x, y) = xy$ and $h(x, y) = |x - y|$, the latter leading to Gini's mean-difference statistic.

## 2. LIMITING DISTRIBUTIONS UNDER $\mathcal{H}_0$

Let $h$ be a symmetric kernel function [i.e., $h(x, y) = h(y, x)$ for all $x, y \in \mathbb{R}$] or let $h$ be antisymmetric [$h(x, y) = -h(y, x)$]. We suppose that

$$\mathcal{E} h^2(X_1, X_2) < \infty. \tag{2.1}$$

Put $\theta = \mathcal{E} h(X_1, X_2)$ ($= 0$ in the antisymmetric case), and define the projection $\tilde{h}$ via

$$\tilde{h}(t) = \mathcal{E}\{h(t, X_1) - \theta\}. \tag{2.2}$$

Furthermore, we require that $h$ be a nondegenerate kernel, i.e.,

$$\mathcal{E}\tilde{h}^2(X_1) = \tilde{\sigma}^2 > 0. \tag{2.3}$$

For the random variables $U_k$ introduced in (1.1) we then have the following representation as a linear combination of nondegenerate one-sample $U$-statistics:

$$U_k = U_n^{(3)} - U_k^{(1)} - U_k^{(2)}, \tag{2.4}$$

where

$$U_n^{(3)} = \sum_{1 \le i < j \le n} h(X_i, X_j) - \binom{n}{2}\theta,$$

$$U_k^{(1)} = \sum_{1 \le i < j \le k} h(X_i, X_j) - \binom{k}{2}\theta,$$

$$U_k^{(2)} = \sum_{k+1 \le i < j \le n} h(X_i, X_j) - \binom{n-k}{2}\theta$$

in case of symmetric $h$, while the constant terms vanish when $h$ is antisymmetric.

THEOREM 1. *Let h be an antisymmetric kernel function.*

(a) *Under the assumptions* (2.1) *and* (2.3) *we have as* $n \longrightarrow \infty$

$$\frac{n^{-3}}{\sigma^2}\int_0^1 U_{[(n+1)t]}^2 \, dt \xrightarrow{\mathcal{D}} \int_0^1 B^2(t)\, dt,$$

where $B = \{B(t) : 0 \le t \le 1\}$ *denotes a Brownian bridge (tied-down Brownian motion).*

(b) *Suppose in addition to* (2.1) *and* (2.3) *we also have*

$$\mathcal{E}\,|h(X_1, X_2)|^{\nu} < \infty \qquad \textit{for some} \quad \nu > 2. \tag{2.5}$$

*Let* $g : (0,1) \longrightarrow (0,\infty)$ *by monotonically nondecreasing near 0, monotonically nonincreasing near 1, and such that* $\lim_{\delta \le t \le 1-\delta} g(t) > 0$ *for all* $\delta \in (0, \frac{1}{2})$. *Then as* $n \longrightarrow \infty$

$$\frac{n^{-3}}{\sigma^2}\int_0^1 \frac{U_{[(n+1)t]}^2}{g(t)}\, dt \xrightarrow{\mathcal{D}} \int_0^1 \frac{B^2(t)}{g(t)}\, dt$$

*if and only if*

$$\int_0^1 \frac{t(1-t)}{g(t)}\, dt = \int_0^1 \frac{\mathcal{V}ar\; B(t)}{g(t)}\, dt < \infty.$$

REMARKS. As pointed out in the introduction, the results of Theorem 1 clearly indicate that only one upper/lower-percentage-point table is necessary per weight function, independently of the chosen antisymmetric kernel $h$ and the underlying distribution $F$ of the data.

The weight functions $g$ are intended to make the corresponding test statistics more sensitive on the tails, i.e., near 0 and 1 [as in general $g(t) \longrightarrow 0$ as $t \longrightarrow 0$ or $t \longrightarrow 1$]. In the proof of Theorem 1(b) we will temporarily assume that $g$ is symmetric about $\frac{1}{2}$. This is done to shorten the proof only and is by no means a necessary restriction.

*Proof of Theorem 1.* Part (a): Janson and Wichura (1983) (see in particular Theorem 2.1) present weak approximations of the $U$-statistics in (2.4) under the conditions (2.1) and (2.3) [for a detailed proof see Huse (1988), Lemma 2.1.6]: as $n \longrightarrow \infty$,

$$\begin{gathered}\max_{1\le k\le n}\left|U_k^{(1)} - \sum_{i=1}^{k}(k-2i+1)\tilde{h}(X_i)\right| = O_P(n),\\ \max_{1\le k\le n}\left|U_k^{(2)} - \sum_{i=k+1}^{n}(n+k-2i+1)\tilde{h}(X_i)\right| = O_P(n),\\ \left|U_n^{(3)} - \sum_{i=1}^{n}(n-2i+1)\tilde{h}(X_i)\right| = O_P(n).\end{gathered} \tag{2.6}$$

The statements in (2.6) could be termed a reduction principle. They reduce the problem of determining the limiting distribution of a linear combination of $U$-statistics to the problem of finding the limiting distribution of a sequence of partial sums of independent identically distributed random variables (namely the projections).

The reduction principle (2.6) can easily be seen to imply (compare also Huse 1988, Corollary 2.1.7): as $n \longrightarrow \infty$,

$$\max_{1\le k\le n}\left|U_k - \left(n\sum_{i=1}^{k}\tilde{h}(X_i) - k\sum_{i=1}^{n}\tilde{h}(X_i)\right)\right| = O_P(n). \tag{2.7}$$

Using (2.7), the equality $a^2 - b^2 = (a-b)^2 + 2b(a-b)$, and the fact that the square integral of a Brownian bridge is almost surely finite, part (a) follows now via Donsker's Theorem as in Theorem 2.1.2 in Csörgő and Révész (1981) from the corresponding sup-norm result [see Theorem 4.1 in Csörgő and Horváth (1988a)]:

$$\sup_{0<t<1}\left|\frac{n^{-\frac{3}{2}}}{\sigma}U_{[(n+1)t]} - B_n(t)\right| = o_P(1) \quad \text{as} \quad n \longrightarrow \infty. \tag{2.8}$$

Here $\{B_n\}_{n\ge1}$ is a sequence of Brownian bridges. We note that without loss of generality we can always work on a large enough probability space which accommodates all processes and random variables in question (cf. DeAcosta 1982).

Part (b), sufficiency: We first split the integral into three parts, integrating

$$\frac{1}{g(t)}\left\{\left(\frac{n^{-\frac{3}{2}}}{\sigma}U_{[(n+1)t]}\right)^2 - B_n^2(t)\right\},$$

where $B_n$ is as in (2.8), successively over

$$(\epsilon, 1-\epsilon), \qquad \left(\frac{1}{n+1}, \epsilon\right) \quad \text{and} \quad \left(1-\epsilon, 1-\frac{1}{n+1}\right).$$

For fixed but arbitrary $\epsilon > 0$, on applying the same arguments as used in part (a), by (2.8) we obtain

$$\left| \int_{\epsilon}^{1-\epsilon} \frac{\left( \frac{n^{-\frac{3}{2}}}{\sigma} U_{[(n+1)t]} \right)^2 - B_n^2(t)}{g(t)} \, dt \right| = o_P(1) \quad \text{as} \quad n \longrightarrow \infty. \tag{2.9}$$

By symmetry we now only need to consider the range $(1/(n+1), \epsilon)$. We will employ

$$\sup_{1/(n+1) \leq t \leq n/(n+1)} \frac{1}{\{t(1-t)\}^{\frac{1}{2}}} \left| \frac{n^{-\frac{3}{2}}}{\sigma} U_{[(n+1)t]} - B_n(t) \right| = O_P(1) \quad \text{as} \quad n \longrightarrow \infty \tag{2.10}$$

(cf. Csörgő and Horváth 1988a, Theorem 4.2) in lieu of (2.8). Equation (2.10) is valid under the additional assumption (2.5). We also need the Cauchy-Schwarz inequality and an application of (19.2) in Shepp (1966) to arrive at

$$\left| \int_{1/(n+1)}^{\epsilon} \frac{\left( \frac{n^{-\frac{3}{2}}}{\sigma} U_{[(n+1)t]} \right)^2 - B_n^2(t)}{g(t)} \, dt \right| = o_P(1) \quad \text{as} \quad n \longrightarrow \infty$$

and then $\epsilon \downarrow 0$ by (2.10). Here the argument taken from Shepp (1966) states that

$$\int_0^{\epsilon} \frac{B^2(t)}{g(t)} \, dt < \infty \quad \text{almost surely} \quad \text{if and only if} \quad \int_0^{\epsilon} \frac{t(1-t)}{g(t)} \, dt < \infty. \tag{2.11}$$

To complete the sufficiency part of this proof we note that

$$\int_0^{1/(n+1)} \frac{B^2(t)}{g(t)} \, dt = \int_{n/(n+1)}^{1} \frac{B^2(t)}{g(t)} \, dt = o_P(1) \quad \text{as} \quad n \longrightarrow \infty$$

holds by (2.11) and by the assumed symmetry of $g$ for any Brownian bridge $B$.

Part (b), necessity: Proof by contradiction provides this part immediately from the statement (19.4) in Shepp (1966):

$$\int_0^1 \frac{t(1-t)}{g(t)} \, dt = \infty \quad \text{implies} \quad \int_0^1 \frac{B^2(t)}{g(t)} \, dt = \infty \quad \text{almost surely.}$$

Q.E.D.

THEOREM 2. *Let $h$ be a symmetric kernel function.*

(a) *Under the assumptions* (2.1) *and* (2.3) *we have*

$$\frac{n^{-3}}{\sigma^2} \int_0^1 U_{[(n+1)t]}^2 \, dt \xrightarrow{\mathcal{D}} \int_0^1 \Gamma^2(t) \, dt \qquad \textit{as} \quad n \longrightarrow \infty,$$

*where the Gaussian process $\Gamma = \{\Gamma(t) : 0 \leq t \leq 1\}$ is defined by $\Gamma(t) = (1-t)W(t) + tW(1) - tW(t)$. Here $W$ denotes a standard Brownian motion (Wiener process).*

(b) *Given the conditions* (2.1), (2.3), (2.5), *and* $g$ *as in* (b) *of Theorem 1, we have as* $n \longrightarrow \infty$

$$\frac{n^{-3}}{\sigma^2}\int_0^1 \frac{U^2_{[(n+1)t]}}{g(t)}\,dt \xrightarrow{\mathcal{D}} \int_0^1 \frac{\Gamma^2(t)}{g(t)}\,dt$$

*if and only if*

$$\int_0^1 \frac{t(1-t)}{g(t)}\,dt = \int_0^1 \frac{\mathcal{V}ar\ \Gamma(t)}{g(t)}\,dt < \infty.$$

REMARK. We note that although the expected value and the variance of $\Gamma(t)$ and of $B(t)$ coincide, the Gaussian process $\Gamma$ is not a Brownian bridge, because $\mathcal{C}ov\left(\Gamma(s),\Gamma(t)\right) = (1-2s)(1-2t)(s\wedge t)+(3-2s-2t) \neq \mathcal{C}ov\left(B(s),B(t)\right) = s\wedge t - st$, where $s \wedge t$ denotes the minimim of $s$ and $t$.

*Proof of Theorem 2.* Instead of (2.6) we use as the starting point here

$$\max_{1\le k\le n}\left|U_k^{(1)} - k\sum_{i=1}^{k}\tilde{h}(X_i)\right| = O_P(n),$$

$$\max_{1\le k\le n}\left|U_k^{(2)} - (n-k)\sum_{i=k+1}^{n}\tilde{h}(X_i)\right| = O_P(n), \tag{2.12}$$

$$\left|U_n^{(3)} - n\sum_{i=1}^{n}\tilde{h}(X_i)\right| = O_P(n)$$

an $n \longrightarrow \infty$. Equation (2.12) is a consequence of Theorem 1 of Hall (1979) and implies immediately

$$\max_{1\le k\le n}\left|U_k - \left\{(n-k)\sum_{i=1}^{n}\tilde{h}(X_i) + k\left(\sum_{i=1}^{n}\tilde{h}(X_i) - \sum_{i=1}^{k}\tilde{h}(X_i)\right)\right\}\right| = O_P(n) \tag{2.13}$$

as $n \longrightarrow \infty$. Comparing (2.13) with the previously obtained (2.7) for the antisymmetric case, we notice that a different Gaussian process emerges as the limit in the symmetric case. This fact is reflected in saying [cf. Theorems 2.1 and 2.2 of Csörgő and Horváth (1988a)] that under the assumptions (2.1) and (2.3) one can define a sequence of Gaussian processes $\{\Gamma_n(t) : 0 \le t \le 1\}_{n\ge 1}$ such that (2.8) holds true with $\Gamma_n$ replacing $B_n$. And we have (2.10) as well this way on assuming also (2.5), where for each $n \ge 1$, $\Gamma_n$ is the same Gaussian process as that of Theorem 2. The remainder of the proof of Theorem 2 follows along the same lines as the proof of Theorem 1 with minor modifications to accommodate the new process $\Gamma$. Q.E.D.

Figure 1 shows three different sample paths of a Brownian bridge $B$, while Figure 2 displays the corresponding sample paths of a $\Gamma$-process. All sample paths were generated from realizations of a Wiener process via $B(t) = W(t) - tW(1)$ and $\Gamma(t) = (1-t)W(t) + tW(1) - tW(t)$. The simulation procedure is explained in Eastwood and Eastwood (1990). As proposed in the introduction, Theorems 1 and 2 give the limiting distributions for any changepoint problem under the assumption of no change in distribution.

In the next section we will suggest methods to tabulate these limiting distributions.

## 3. TABULATING THE DISTRIBUTION OF GAUSSIAN FUNCTIONALS

The major goal of the present section is to find upper percentage points for

$$\int_0^1 B^2(t)\,dt, \qquad \int_0^1 \frac{B^2(t)}{g(t)}\,dt, \qquad \int_0^1 \Gamma^2(t)\,dt, \quad \text{and} \quad \int_0^1 \frac{\Gamma^2(t)}{g(t)}\,dt,$$

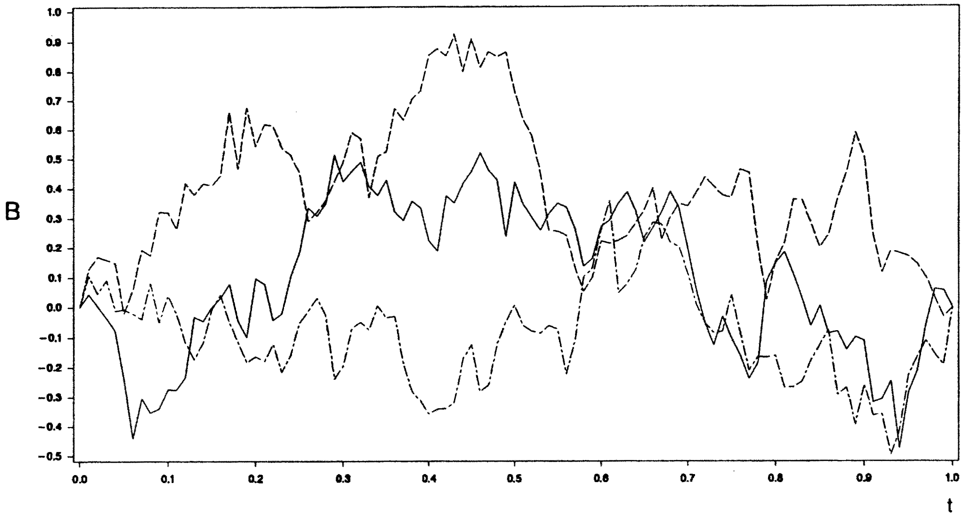


FIGURE 1: Three sample Brownian bridge processes approximated using 101 equally spaced sample points.

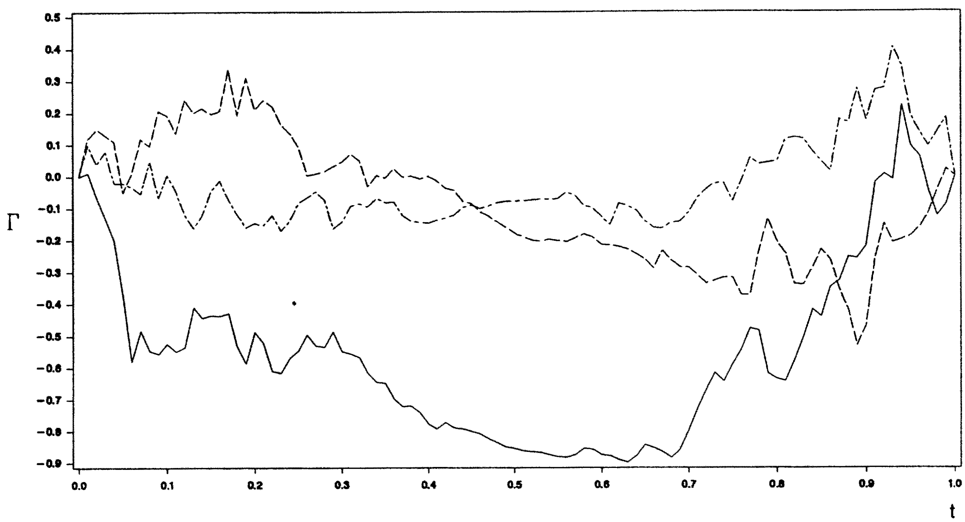


FIGURE 2: Three sample gamma processes approximated using 101 equally spaced sample points.

or, more generally speaking, for $\int_0^1 X^2(t)\,dt$, where $X$ denotes an arbitrary Gaussian process. Nonnumerical methods already established in the literature rely mostly on an application of Mercer's Theorem (see, e.g., Shorack and Wellner 1986, Chapter 5) and on the computation of the eigenvalues of certain differential equations.

These methods, however, have been shown to fail often, especially when $X \neq B$, $W$ or when weight functions of a slightly more complicated nature are involved. For example, taking $X(t) = \sqrt{\psi(t)}B(t)$ leads to a Riccati differential equation which will have closed-form solutions (represented by Bessel functions) at most for weight functions of the form $\psi(t) = ct^{\alpha}$, $\alpha > -2$ (cf. Kamke 1951, A.48).

MacNeil (1974) used a similar approach to solve the problem for $X^2(t) = at^{\alpha}W^2(t)$

when $\alpha > -2$. The condition $\alpha > -2$ results from the requirement that the square integral be finite almost surely. A detailed discussion of the merits of previous attempts to tabulate the distribution of square integrals of Gaussian functionals is given in Chapter 3 of Huse (1988).

In order to find $P(\int_0^1 X^2(t)\,dt \leq z)$ for given $z > 0$ and conversely $z$ for given probability level $\alpha$ for a large enough class of Gaussian functionals, we employ the following method, which for simplicity's sake is introduced here using the example $X = \Gamma$ (see also Eastwood and Eastwood 1990). The starting point is the classical Paley-Wiener representation (1934) of Brownian motion [see, e.g., Csörgő and Révész (1981), formula (1.8.4)]:

$$W(t) = Y_0 t + \sqrt{2} \sum_{k=1}^{\infty} \frac{\sin k\pi t}{k\pi} Y_k, \qquad 0 \leq t \leq 1. \tag{3.1}$$

The convergence of the infinite series in (3.1) is uniform and absolute in the unit interval, and $\{Y_k\}_{k \geq 0}$ denotes a sequence of independent identically distributed standard normal random variables.

As $\Gamma^2(t) = \{(1-t)W(t) + tW(1) - tW(t)\}^2$ for $t \in [0,1]$, Equation (3.1) and the interchange of integration and summation imply

$$\begin{aligned} \int_0^1 \Gamma^2(t)\,dt &= 4Y_0^2 \int_0^1 t^2(1-t)^2\,dt \\ &\quad + 4\sqrt{2} Y_0 \sum_{k=1}^{\infty} \frac{Y_k}{k\pi} \int_0^1 t(1-t)(1-2t) \sin k\pi t\,dt \\ &\quad + 2 \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{Y_k}{k\pi} \frac{Y_m}{m\pi} \int_0^1 (1-2)^2 \sin k\pi t \sin m\pi t\,dt \\ &= \lim_{K \to \infty} \sum_{k,m=0}^{K} a_{km} Y_k Y_m, \end{aligned}$$

where

$$a_{km} = \begin{cases} \dfrac{2}{15} & \text{if } k = m = 0, \\ \dfrac{1}{3(k\pi)^2} - \dfrac{2}{(k\pi)^4} & \text{if } k = m \geq 1, \\ \dfrac{24\sqrt{2}}{(k\pi)^4} & \text{if } k \geq 1,\ k \text{ even},\ m = 0, \\ \dfrac{24\sqrt{2}}{(m\pi)^4} & \text{if } m \geq 1,\ m \text{ even},\ k = 0, \\ \dfrac{32}{\{(k^2 - m^2)\pi^2\}^2} & \text{if } k, m \geq 1,\ k \neq m,\ k + m \text{ even}, \\ 0 & \text{otherwise}. \end{cases} \tag{3.2}$$

This leads us to the following general observations: if $T_X = \int_0^1 X^2(t)\,dt$ is any given Gaussian functional, then it is an immediate consequence of the Paley-Wiener representation (3.1) that there exist real numbers $a_{km}$, $k, m \geq 0$, and a sequence $\{Y_k\}_{k \geq 0}$ of independent identically distributed $\mathsf{N}(0,1)$ random variables such that

$$P\left(\int_0^1 X^2(t)\,dt \leq z\right) \simeq P\left(\sum_{k,m=0}^{K} a_{km} Y_k Y_m \leq z\right). \tag{3.3}$$

TABLE 1

| | Cutoff points for | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $P\left(\int_0^1 tW^2(t)\,dt \le z\right) = \alpha$ | | | | $P\left(\int_0^1 t^2W^2(t)\,dt \le z\right) = \alpha$ | | | |
| $\alpha$ | $z_{50}$ | $z_{100}$ | $z_{\infty}$ | $z_{\alpha}$ | $z_{50}$ | $z_{100}$ | $z_{\infty}$ | $z_{\alpha}$ |
| 0.010 | 0.0407 | 0.0412 | 0.0417 | 0.01685 | 0.0244 | 0.0247 | 0.0250 | 0.0101 |
| 0.025 | 0.0410 | 0.0415 | 0.0420 | 0.02215 | 0.0246 | 0.0249 | 0.0252 | 0.0135 |
| 0.050 | 0.0419 | 0.0424 | 0.0429 | 0.0288 | 0.0253 | 0.0256 | 0.0259 | 0.0178 |
| 0.100 | 0.0457 | 0.0462 | 0.0467 | 0.0402 | 0.0281 | 0.0284 | 0.0287 | 0.0254 |
| 0.500 | 0.1756 | 0.1763 | 0.1768 | 0.18105 | 0.1278 | 0.1282 | 0.1285 | 0.1305 |
| 0.900 | 0.8264 | 0.8269 | 0.8274 | 0.82545 | 0.6315 | 0.6318 | 0.6322 | 0.6313 |
| 0.950 | 1.153 | 1.153 | 1.154 | 1.152 | 0.8848 | 0.8852 | 0.8855 | 0.8846 |
| 0.975 | 1.492 | 1.493 | 1.493 | 1.4921 | 1.148 | 1.149 | 1.149 | 1.148 |
| 0.990 | 1.954 | 1.955 | 1.955 | 1.9552 | 1.507 | 1.507 | 1.508 | 1.508 |

Furthermore, after diagonalizing the matrix $A = (a_{km})$, we obtain that

$$P\left(\int_0^1 X^2(t)\,dt \le z\right) \simeq P\left(\sum_{k=0}^{K} \lambda_k Z_k^2 \le z\right) \tag{3.4}$$

for appropriate $\lambda_0 > \lambda_1 > \cdots > \lambda_K > 0$ and $K$ chosen suitably large. In other words, the distribution of the Gaussian functional $T_X$ can be approximated arbitrarily closely (for $K$ extremely large) by a quadratic form in independent $\mathsf{N}(0,1)$ random variables without having to solve any differential equations. This approach appears to have been overlooked thus far. It works even in cases of complicated, noncontinuous covariance functions and leads to acceptable results in all known cases, as is demonstrated next. Applying the previously described method, we obtain for $\alpha \ge 1$ even:

$$\begin{aligned}\int_0^1 t^\alpha W^2(t)\,dt = {}& \int_0^1 t^\alpha B^2(t)\,dt + \frac{1}{\alpha+3}\,Y_0^2 \\ & + 2^{\frac{3}{2}} \sum_{k=1}^{\infty} \frac{Y_0 Y_k}{(k\pi)^2} (-1)^k (\alpha+1)! \sum_{n=0}^{\alpha/2} \frac{(-1)^{n+1}}{(\alpha+1-2n)!(k\pi)^{2n}},\end{aligned}$$

and for $\alpha \ge 1$ odd:

$$\begin{aligned}\int_0^1 t^\alpha W^2(t)\,dt = {}& \int_0^1 t^\alpha B^2(t)\,dt + \frac{1}{\alpha+3}\,Y_0^2 \\ & + 2^{\frac{3}{2}} \sum_{k=1}^{\infty} \frac{Y_0 Y_k}{(k\pi)^2} (-1)^k (\alpha+1)! \sum_{n=0}^{(\alpha+1)/2} \frac{(-1)^{n+1}}{(\alpha+1-2n)!(k\pi)^{2n}} \\ & + 2^{\frac{3}{2}} \sum_{k=1}^{\infty} \frac{Y_0 Y_k}{(k\pi)^{\alpha+3}} (-1)^{(\alpha+1)/2} (\alpha+1)!.\end{aligned}$$

The same method can then be used to obtain representations for $\int_0^1 t^\alpha B^2(t)\,dt$ (cf. Eastwood and Eastwood 1990, Section 6). Here we are interested in the special cases $\alpha = 1, 2$, which were already studied by MacNeill (1974). MacNeill's values are quoted as $z_\alpha$ in Table 1 and are used to demonstrate the exactness of our method. We calculated:

$$\int_0^1 tB^2(t)\,dt = \frac{1}{2}\sum_{k=1}^{\infty}\frac{Y_k^2}{(k\pi)^2} - 8\sum_{\substack{k=1\\ k\neq m,}}^{\infty}\sum_{\substack{m=1\\ k+m \text{ odd}}}^{\infty}\frac{Y_kY_m}{(k^2-m^2)^2\pi^4}$$

for $\alpha = 1$ and

$$\int_0^1 t^2B^2(t)\,dt = \frac{1}{3}\sum_{k=1}^{\infty}\frac{Y_k^2}{(k\pi)^2} - \frac{1}{2}\sum_{k=1}^{\infty}\frac{Y_k^2}{(k\pi)^4} + 8\sum_{\substack{k=1\\ k\neq m}}^{\infty}\sum_{m=1}^{\infty}(-1)^{k+m}\frac{Y_kY_m}{(k^2-m^2)^2\pi^4}$$

for $\alpha = 2$. Combining the last four formulae, we computed the $K \times K$ matrix $A$ as required for (3.3). To obtain Table 1 from here, the following steps were performed for $K = 50, 60, \ldots, 100$:

*Step 1.* Enter $A$ in the computer (we used a VAX 8800, FORTRAN routines, 64-bit double precision). Call IMSL library subroutine DEVCSF to compute all eigenvalues of the real symmetric matrix $A$. Note that, using Lemma 3.1.3 in Huse (1988) and Mercer's Theorem, it can be shown that $A$ is a positive definite matrix and that consequently all eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_K$ must be positive. The subroutine DEVCSF reports the $\lambda_i$'s in ascending order. As a result we now have

$$P\left(\int_0^1 X^2(t)\,dt \le z\right) \simeq P\left(\sum_{k,m=1}^{K} a_{km}Y_kY_m \le z\right) = P\left(\sum_{k=1}^{K}\lambda_k Z_k^2 \le z\right).$$

*Step 2.* For given value of the parameter $K$, compute the first three cumulants of the last quadratic form as

$$c_1^* = \sum_{k=1}^{K}\lambda_k, \qquad c_2^* = 2\sum_{k=1}^{K}\lambda_k^2, \quad \text{and} \quad c_3^* = 8\sum_{k=1}^{K}\lambda_k^3.$$

We denote these by $c_i^* = c_{i,K}^*$ for $i = 1, 2, 3$. Whenever possible we also compute

$$\mathcal{E}\int_0^1 X^2(t)\,dt = c_{1,\infty}, \qquad \mathcal{V}ar\int_0^1 X^2(t)\,dt = c_{2,\infty}, \quad \text{and} \quad c_{3,\infty}$$

using the recursively defined covariance integral of Shorack and Wellner (1986, pp. 212–213). Next renormalize the cumulants to arrive at

$$c_{1,K} = c_{1,K}^*, \qquad c_{2,K} = \frac{c_{2,K}^*}{2}, \qquad c_{3,K} = \frac{c_{3,K}^*}{8}, \quad \text{and} \quad h' = \frac{c_{2,K}^3}{c_{3,K}^2}.$$

*Step 3.* By Imhof (1961),

$$P\left(\int_0^1 X^2(t)\,dt \le z\right) \simeq P\left(\sum_{k=1}^{K}\lambda_k Z_k^2 \le z\right) \simeq P(\chi_{h'}^2 \le y),$$

where $\chi_{h'}^2$ denotes a (central) chi-square random variable with $h'$ degrees of freedom and where $z$ and $y$ are related by the equation $y = (z - c_{1,K})(h'/c_{2,K})^{\frac{1}{2}} + h'$. For various levels of $\alpha$ we used the IMSL routine DCHIIN to evaluate the inverse of the appropriate chi-square

TABLE 2

| | Cutoff points for $P\left(\int_0^1 \Gamma^2(t)\,dt \le z\right) = \alpha$ | | | | |
|---|---|---|---|---|---|
| $\alpha$ | $z_{50}$ | $z_{70}$ | $z_{90}$ | $z_{100}$ | $\hat{z}$ |
| 0.010 | 0.0115 | 0.0116 | 0.0117 | 0.0117 | 0.00037 |
| 0.025 | 0.0117 | 0.0118 | 0.0118 | 0.0119 | 0.00129 |
| 0.050 | 0.0121 | 0.0122 | 0.0123 | 0.0123 | 0.00335 |
| 0.100 | 0.0140 | 0.0141 | 0.0142 | 0.0142 | 0.00879 |
| 0.250 | 0.0274 | 0.0275 | 0.0276 | 0.0276 | 0.03278 |
| 0.500 | 0.0823 | 0.0824 | 0.0825 | 0.0825 | 0.09935 |
| 0.750 | 0.2165 | 0.2166 | 0.2167 | 0.2167 | 0.22945 |
| 0.900 | 0.4297 | 0.4298 | 0.4299 | 0.4299 | 0.41435 |
| 0.950 | 0.6048 | 0.6049 | 0.6049 | 0.6050 | 0.55900 |
| 0.975 | 0.7869 | 0.7870 | 0.7871 | 0.7871 | 0.70612 |
| 0.990 | 1.035 | 1.035 | 1.035 | 1.035 | 0.90314 |

distribution function. The values $y = y_{\alpha,K}$ found through DCHIIN were then converted to values $z_K$ via the above equation. Values of $z_K$ for $K = 50$ and 100 are reported in Table 1. Values of $z_\infty$ are based on $c_{i,\infty}$ of step 2. A more extensive collection of tables is in Eastwood and Eastwood (1990). Table 2 displays the cutoff points for the square integral of the $\Gamma$-process. Using Monte Carlo simulations, the distribution of sup-norm functionals of the $\Gamma$-process has also been tabulated (cf. Eastwood and Eastwood 1990).

## 4. CONSISTENCY

In order to verify the consistency of our procedure we need the following notation:

$$\begin{aligned}\theta_1 &= \mathcal{E}\, h(X_{[n\lambda]-1}, X_{[n\lambda]}),\\ \theta_2 &= \mathcal{E}\, h(X_{[n\lambda]+1}, X_{[n\lambda]+2}),\\ \theta_3 &= \mathcal{E}\, h(X_{[n\lambda]}, X_{[n\lambda]+1})\end{aligned} \tag{4.1}$$

for $\lambda \in (0,1)$ fixed. Assuming $\mathcal{H}_A$ to be true, $\theta_1$ denotes the expected value before the change, $\theta_2$ the expected value after the change, and $\theta_3$ the expected value around the changepoint. Note that in the case of an antisymmetric kernel $h$, $\theta_1 = \theta_2 = 0$. Keeping this in mind, we formulate

THEOREM 3. *Assume that $\mathcal{H}_A$ is true, Put $\log^+(x) = \log(\max\{x, 1\})$. Suppose that*

$$\begin{aligned}\mathcal{E}\left|h(X_{[n\lambda]-1}, X_{[n\lambda]})\right| &< \infty,\\ \mathcal{E}\left|h(X_{[n\lambda]+1}, X_{[n\lambda]+2})\right| &< \infty,\\ \mathcal{E}\left\{\left|h(X_{[n\lambda]}, X_{[n\lambda]+1})\right|\log^+\left(\left|h(X_{[n\lambda]}, X_{[n\lambda]+1})\right|\right)\right\} &< \infty.\end{aligned}$$

(a) *Let $h$ be a symmetric kernel. Then*

$$\lim_{n\to\infty} \frac{1}{n^2}\, U_{[(n+1)t]} = \begin{cases} (\lambda - t)t\theta_1 + (1-\lambda)t\theta_3 & \text{if } \ 0 < t \le \lambda, \\ (t-\lambda)(1-t)\theta_2 + \lambda(1-t)\theta_3 & \text{if } \ \lambda \le t < 1 \end{cases}$$

*in probability.*

(b) *Let h be an antisymmetric kernel. Then*

$$\lim_{n\to\infty} \frac{1}{n^2} U_{[(n+1)t]} = \begin{cases} (1-\lambda)t\theta_3 & \text{if} \quad 0 < t \le \lambda, \\ \lambda(1-t)\theta_3 & \text{if} \quad \lambda \le t < 1 \end{cases}$$

*in probability.*

A first proof of Theorem 3(a) was presented by Csörgő and Horváth (1988a). A more detailed version, using the weak-convergence results on two-sample $U$-statistics due to P.K. Sen (1974, 1977), was given in Huse (1988). The proof of (b) requires only minor modifications, which were presented in Huse (1988).

As it is a direct consequence of Theorems 1 and 2, respectively, that $\lim_{n\to\infty} (1/n^2)\times U_{[(n+1)t]} = 0$ in probability under the null hypothesis of no change, Theorem 3 immediately implies consistency under any AMOC alternative that satisfies the condition that at least one of $\theta_1$, $\theta_2$, and $\theta_3$ is nonzero.

## ACKNOWLEDGEMENT

This article is partially based on the author's Ph.D. thesis written at Carleton University under the supervision of Miklós Csörgő. The author wishes to thank Miklós Csörgő for his dedication, numerous discussions, and advice freely given during the time in which the thesis was written and also since then.

## REFERENCES

Csörgő, M., and Horváth, L. (1988a). Invariance principles for changepoint problems, *J. Multivariate Anal.*, 27, 151–168.

Csörgő, M., and Horváth, L. (1988b). Nonparametric methods for changepoint problems. *Handbook of Statistics 7 (Quality Control and Reliability)* (P.R. Krishnaiah and C.R. Rao, *eds.*), Elsevier Science, 1988, 403–425.

Csörgő, M., and Révész, P. (1981). *Strong Approximations in Probability and Statistics.* Academic Press, New York.

DeAcosta, A. (1982). Invariance principles in probability for triangular arrays of $B$-valued random vectors and some applications. *Ann. Probab.*, 10, 346–373.

Eastwood, B., and Eastwood, V. (1990). On the distribution of some Gaussian functionals related to changepoint problems. Technical Report 146, Laboratory for Research in Statistics and Probability, Carleton University, Ottawa.

Eastwood, V. (1990). Some recent developments concerning asymptotic distributions of pontograms. *Math. Proc. Cambridge Philos. Soc.*, 108, 559–567.

Grad, A., and Solomon, H. (1955). Distribution of quadratic forms and some applications. *Ann. Math. Statist.*, 26, 464–477.

Hall, P. (1979). On the invariance principle for $U$-statistics. *Stochastic Process. Appl.*, 9, 163–174.

Huse, V. (1988). On some nonparametric methods for changepoint problems. Ph.D. Thesis, Carleton University, Ottawa.

Huse, V. (1989). Asymptotic properties for the sequential CUSUM procedure. *Statist. Probab. Lett.*, 7, 73–80.

Huse-Eastwood, V. (1989). Limiting distributions of Poisson bridges, Kendall-Kendall pontograms and related processes. *Limit Theorems in Probability and Statistics, Colloq. Math. Soc.* János Bolyai 57, Pécs, Hungary, 271–279.

Imhof, J.P. (1961). Computing the distribution of quadratic forms in normal variables, *Biometrika*, 48, 419–426.

Janson, S., and Wichura, M. (1983). Invariance principles for stochastic area and related stochastic integrals, *Stochastic Process. Appl.*, 16, 71–84.

Kamke, E. (1951). *Differentialgleichungen, Lösungsmethoden und Lösungen*, Band 1, 4. verb. Auflage, Akadem. Verlagsges., Leipzig.

MacNeill, I. (1974). Tests for change of parameter at unknown times and distributions of some related functionals on Brownian motion. *Ann. Statist.*, 2, 950–962.

Pettitt, A.N. (1979). A nonparametric approach to the changepoint problem. *Appl. Statist.*, 28, 126–135.

Sen, P.K. (1974). Weak convergence of generalized $U$-statistics. *Ann. Probab.*, 2, 90–102.

Sen, P.K. (1977). Almost sure convergence of generalized $U$-statistics. *Ann. Probab.*, 5 287–290.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.
Shepp, L. (1966). Radon-Nikodym derivatives of Gaussian measures. *Ann. Math. Statist.*, 37, 321–354.
Shorack, G., and Wellner, J. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.
Wolfe, D., and Schechtman, E. (1984). Nonparametric statistical procedures for the changepoint problem. *J. Statist. Plann. Inference*, 9, 389–396.

*Received 28 May 1990*
*Revised 20 July 1992*
*Accepted 2 February 1993*

*Department of Mathematics and Statistics*
*Academia University*
*Wolfville, Nova Scotia*
*Canada B0P 1X0*

# LINKED CITATIONS





**Some Nonparametric Methods for Changepoint Problems**

Vera R. Eastwood

*The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, Vol. 21, No. 2. (Jun., 1993), pp. 209-222.

Stable URL:

http://links.jstor.org/sici?sici=0319-5724%28199306%2921%3A2%3C209%3ASNMFCP%3E2.0.CO%3B2-W



## REFERENCES

**Invariance Principles in Probability for Triangular Arrays of B-Valued Random Vectors and Some Applications**

Alejandro de Acosta

*The Annals of Probability*, Vol. 10, No. 2. (May, 1982), pp. 346-373.

Stable URL:

http://links.jstor.org/sici?sici=0091-1798%28198205%2910%3A2%3C346%3AIPIPFT%3E2.0.CO%3B2-8

**Distribution of Quadratic Forms and Some Applications**

Arthur Grad; Herbert Solomon

*The Annals of Mathematical Statistics*, Vol. 26, No. 3. (Sep., 1955), pp. 464-477.

Stable URL:

http://links.jstor.org/sici?sici=0003-4851%28195509%2926%3A3%3C464%3ADOQFAS%3E2.0.CO%3B2-P

**Computing the Distribution of Quadratic Forms in Normal Variables**

J. P. Imhof

*Biometrika*, Vol. 48, No. 3/4. (Dec., 1961), pp. 419-426.

Stable URL:

http://links.jstor.org/sici?sici=0006-3444%28196112%2948%3A3%2F4%3C419%3ACTDOQF%3E2.0.CO%3B2-D

# LINKED CITATIONS




**Tests for Change of Parameter at Unknown Times and Distributions of Some Related Functionals on Brownian Motion**

Ian B. MacNeill

*The Annals of Statistics*, Vol. 2, No. 5. (Sep., 1974), pp. 950-962.

Stable URL:

http://links.jstor.org/sici?sici=0090-5364%28197409%292%3A5%3C950%3ATFCOPA%3E2.0.CO%3B2-F

**A Non-Parametric Approach to the Change-Point Problem**

A. N. Pettitt

*Applied Statistics*, Vol. 28, No. 2. (1979), pp. 126-135.

Stable URL:

http://links.jstor.org/sici?sici=0035-9254%281979%2928%3A2%3C126%3AANATTC%3E2.0.CO%3B2-1

**Weak Convergence of Generalized U-Statistics**

Pranab Kumar Sen

*The Annals of Probability*, Vol. 2, No. 1. (Feb., 1974), pp. 90-102.

Stable URL:

http://links.jstor.org/sici?sici=0091-1798%28197402%292%3A1%3C90%3AWCOG%3E2.0.CO%3B2-Z

**Almost Sure Convergence of Generalized U-Statistics**

Pranab Kumar Sen

*The Annals of Probability*, Vol. 5, No. 2. (Apr., 1977), pp. 287-290.

Stable URL:

http://links.jstor.org/sici?sici=0091-1798%28197704%295%3A2%3C287%3AASCOG%3E2.0.CO%3B2-9

**Radon-Nikodym Derivatives of Gaussian Measures**

L. A. Shepp

*The Annals of Mathematical Statistics*, Vol. 37, No. 2. (Apr., 1966), pp. 321-354.

Stable URL:

http://links.jstor.org/sici?sici=0003-4851%28196604%2937%3A2%3C321%3ARDOGM%3E2.0.CO%3B2-O